RESEARCH AND DEVELOPMENT TECHNICAL REPORT
CECOM-TR-96-1

A COMPARISON OF EXPONENTIAL SMOOTHING AND TIME SERIES MODELS FOR FORECASTING CONTRACT ESTIMATES-AT-COMPLETION

RICHARD J. D'ACCARDI, Ph.D.
COST ANALYSIS DIVISION
DIRECTORATE OF RESOURCE MANAGEMENT

February 1996

DISTRIBUTION STATEMENT

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED

CECOM
U.S. ARMY COMMUNICATIONS-ELECTRONICS COMMAND
DIRECTORATE OF RESOURCE MANAGEMENT
FORT MONMOUTH, NEW JERSEY 07703-5034

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

The citation of trade names and names of manufacturers in this report is not to be construed as official Government endorsement or approval of commercial products or services referenced herein.

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1996	Technical Report: 1995

4. TITLE AND SUBTITLE
A COMPARISON OF EXPONENTIAL SMOOTHING AND TIME SERIES MODELS FOR FORECASTING CONTRACT ESTIMATES-AT-COMPLETION

5. FUNDING NUMBERS

6. AUTHOR(S)
Richard J. D'Accardi, Ph.D.

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)
US Army Communications-Electronics Command (CECOM)
Directorate of Resource Management
Cost Analysis Division, ATTN: AMSEL-CP-CA-SV
Fort Monmouth, NJ 07703-5034

8. PERFORMING ORGANIZATION REPORT NUMBER
CECOM-TR-96-1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Several models are embedded in Performance Analyzer (PA) that provide reasonable forecasts of program cost using information contained in contractor-provided monthly Cost Performance Reports (CPR). Considering the nature of this data, both Exponential Smoothing and Time Series models are good candidates to forecast Estimates-at-Completion (EAC). The thrust of this report is twofold. First, to expand the currently available forecasting models that are components of PA to include Time Series and Exponential Smoothing models. Second, for known contract cost performance data, using PA generated forecasts as a benchmark, a comparison of the models is made to highlight the differences between them and to provide a basis for choosing the best model for determining EAC's.

14. SUBJECT TERMS
Exponential Smoothing; Exponential Smoothing models; Time Series; Time Series models; Estimates-at-Completion; Performance Analyzer; Forecasting; Cumulative ACWP; Actual Cost of Work Performed

15. NUMBER IF PAGES
25

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

CONTENTS

FIGURES

FIGURES (Continued)

TABLES

A COMPARISON OF EXPONENTIAL SMOOTHING AND TIME SERIES MODELS FOR FORECASTING CONTRACT ESTIMATES-AT-COMPLETION

by

Richard J. D'Accardi, Ph.D.
Cost Analysis Division
Directorate of Resource Management
CECOM
Fort Monmouth, NJ

1.0 INTRODUCTION

Several models are embedded in Performance Analyzer (PA) that provide reasonable forecasts of program cost using information contained in monthly Cost Performance Reports (CPR). These models require at least three to six months of data to generate reasonable estimates. However, if a program is rebaselined or exhibits anomalous cost efficiency, these models may give misleading results. At best, if a contract is rebaselined, at least three months of data will have to be accumulated to again achieve reasonable forecasts. Therefore, additional models may be appropriate to provide another decision theoretic for analysts to follow that will compensate for cost anomalies. This paper addresses the use of either Time Series or Exponential Smoothing models to determine contract Estimates-at-Completion (EAC).

The random nature of the cumulative Actual Cost of Work Performed (ACWP) and the time-dependence of contract expenditures indicate that time-series and exponential smoothing models are good candidates to forecast EACs. Considering the premise that the pattern of a contractor's expenditures becomes characteristic as the contract proceeds, the application of these models should be to characterize the cumulative ACWP. This is best done with as many data points as possible which implies that the models should not be used early on or for contracts of short duration.

The thrust of this paper is twofold. First, to expand the currently available forecasting models that are components of PA to include time series and exponential smoothing models. Second, for known cost performance data, using PA generated forecasts as a benchmark, a comparison of the models is made to highlight the differences between them and to provide a basis for choosing the best model for determining EACs.

Time Series and Exponential Smoothing forecasting models are treated as compliments to the models incorporated in Performance Analyzer. Because of their ability to track anomalies, within reason, it is recommended that they be used for contracts that display large changes in cost efficiency or are, in fact, problem children.

2.0 SUMMARY OF THE MODELS

Time series are either stationary (i.e., statistical properties do not change with respect to time) or non-stationary (i.e., not in statistical equilibrium). Non-stationary data can be adjusted so that existing techniques of stationary time series can be applied. Smoothing techniques [1], [2],[5], provide a means of removing or reducing volatile short term fluctuations (non-stationarities) in time series. Exponential smoothing is the technique that is discussed in subsequent sections. Difference filtering can also be used to remove non-stationarities in the data rendering the series stationary. Then, once a model for the stationary series is obtained, a backward filter is applied to the fitted stationary model so that future values of the observed series can be forecast [1], [3].

2.1 Time Series Models

Time series that occur in practice are usually non-stationary and require *adjustments* before the existing techniques of stationary time series analysis can be applied [1]. Adjustments are usually accomplished by applying a filter to the observed non-stationary data. The filter is a mathematical function that will remove non-stationary components of the series, transforming it into a stationary time series. The most often used method is difference filtering. A first order difference equation is defined by [1], [3]:

$$y_t = x_t - x_{t-1} \tag{2.1}$$

where x is the observed non-stationary time series and y is the first differenced series. A second order difference equation is defined as:

$$w_t = x_t - 2x_{t-1} + x_{t-2} \tag{2.2}$$

where w is the second differenced series. Either of these will usually be sufficient to remove the non-stationarities from most practically occurring non-stationary time series.

To determine whether or not the observed series exhibits stationary or non-stationary properties, the sample autocorrelation function of the data and a trend test (e.g., Kendall's Tau) are used [3], [4]. Once the data have been rendered stationary, the model is fitted and a *backwards* filter (Equations 2.1 or 2.2 as appropriate) is inserted into the model for simulation of the original series and for forecasting future values of the series.

In order to forecast values for an observed series, parametric time series models either autoregressive, moving average, or a combination of the two are used. These stationary stochastic models assume the process (series) remains in equilibrium about a constant mean. The general autoregressive model [1] is given by:

$$x_t - \mu = \alpha_1 (x_{t-1} - \mu) + \cdots + \alpha_m (x_{t-m} - \mu) + Z_t \tag{2.3}$$

where μ is the mean of x_t, Z_t is a purely random process [1], and m is the order of the process. The general moving-average process is given by:

$$x_t - \mu = Z_t - \beta_1 Z_{t-1} - \cdots - \beta_q Z_{t-q} \tag{2.4}$$

where μ and Z are as defined in equation 2.1, and q is the order of the process. The general mixed autoregressive-moving averages process is given by:

$$x_t - \mu = \alpha_1 (x_{t-1} - \mu) + \cdots + \alpha_m (x_{t-m} - \mu) + Z_t - \beta_1 Z_{t-1} - \cdots - \beta_q Z_{t-q} \tag{2.5}$$

where q is independent of m.

The criterion for selecting the order of the process which will give the best fit is the residual variance for the different orders of the models fitted to the data. The residual variances for the autoregressive, moving averages and mixed processes are plotted against the order of the process. The minimum residual variance will correspond to the process (and its order) that will give the best fit to the data [3]. Once this is done, an appropriate *backwards* filter is applied to obtain forecasting models of the form [1]:

$$\hat{x}_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \phi_3 x_{t-3} + \cdots + \phi_n x_{t-n} \tag{2.6}$$

where ϕ is determined from the model parameters and $\{x_{t-n}\}$ are historical values of the series at time *t*.

2.2 Exponential Smoothing Models

The accuracy of a forecast is influenced by the observed data and the form of the model chosen to represent the underlying process. One of the simplest ways to smooth a series of observations is to use an n period moving average model [5] of the form:

$$M_t = \left(\frac{x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-n}}{n} \right) \tag{2.7}$$

where x_t is the original series and M_t is the smoothed series. This process uses past values and a present value to obtain each M_t.

Exponential smoothing involves the use of an n period exponentially weighted moving average model. In exponential smoothing, the values of the model coefficients are computed such that the sum of the squares of the residuals (differences) between the *actuals* and values forecast by the model are minimized. Also, data obtained t periods ago is weighted by $\beta^t = (1-\alpha)^t$ where $\beta < 1.0$. The term α is the smoothing factor. Therefore, data obtained in the past (perhaps near the start of the process) are weighted very little in the modeling procedure compared to the most recent data obtained. An exponentially smoothed series of an n period process is of the form [2]:

$$\xi_t(t) = \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \cdots + \alpha(1-\alpha)^n x_{t-n} \tag{2.8}$$

and extends over all the data points obtained through t from the origin t-n. As α approaches 1.0, it implies that the most current data are weighted more than for values of α<<1.0, e.g., 0.25. The function ξ_t is a linear combination of all past data where the weight given to the previous data decreases geometrically with *age* [2], [5]. For example, if α=0.25, x_t has a weight of 0.25. The previous values follow the relationship $\alpha(1-\alpha)^k$, and are 0.1875 for x_{t-1}, 0.1406 for x_{t-2}, and so on. Figure 2.1 illustrates this process.

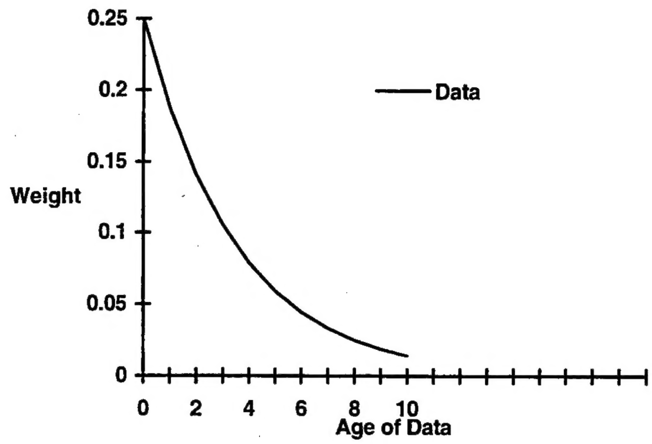

Figure 2.1 Exponential Weighting of Data

When the smoothing constant is large, ξ_t will respond rapidly to changes in the data pattern. The total weight given to the N most recent data [5] is:

$$\alpha \sum_{j=1}^{N-1} \beta^j = 1 - \beta^N, \text{ where } \beta = 1 - \alpha \tag{2.9}$$

β^N approaches $1/e^2$=0.1353 for N>10, hence , about 87% of the weight is given to the N most recent data.

An ℓ steps ahead forecast can be obtained from:

$$\hat{x}_t = \xi_t + \ell r_t \tag{2.10}$$

The forecast, $\hat{x}_t$, is the most recent smoothed value, ξ_t , added to the expected ℓr_t based on the long term trend where r_t is the smoothed series representing the average increase (decrease) in the smoothed series ξ_t [2].

If a sequence of data $\{x(t) = x(1), x(2), \cdots, x(T)\}$ is obtained from a process, ξ(t), the process that characterizes ξ(t) can be expressed as:

$$\xi(t + \tau) = \hat{a}(t) f(\tau) \tag{2.11}$$

where $f(\tau)$ are known fitting functions. For specific characterizations (models), the coefficients $a(t)$ are estimated from the data.

Brown [5] lists ten exponential smoothing models in which the fitting functions include polynomials, exponentials, sinusoids and sums and products of these functions. Of those listed, the linear process:

$$\xi(t) = a_1 + a_2 t \tag{2.12}$$

the quadratic process:

$$\xi(t) = a_1 + a_2 t + \frac{1}{2} a_3 (t-1) \tag{2.13}$$

the simple twelve point sine:

$$\xi(t) = a_1 + a_2 \sin\frac{2\pi t}{12} + a_3 \cos\frac{2\pi t}{12} \tag{2.14}$$

the linear trend with simple sine wave:

$$\xi(t) = a_1 + a_2 t + a_3 \sin\frac{2\pi t}{12} + a_4 \sin\frac{2\pi t}{12} \tag{2.15}$$

seem to have characteristics that fit the underlying process of the CPR data. For more details on the derivations of these models, refer to Brown [5] and Leavenbach [2].

2.3 Performance Analyzer

The generic forecasting equation in Performance Analyzer (PA) is:

$$CAC = \frac{BCWR}{PF} + CumulativeACWP \tag{2.16}$$

where: CAC= Cost-at-Complete
BCWR= Budgeted Cost of Work Remaining
ACWP= Actual Cost of Work Performed
PF= Performance Factor

The forecasting methods in PA that are used as a benchmark for comparing the Time-Series and Exponential Smoothing CACs are a function of how PF is calculated. For the Three-Month Average, the three-month Cost Performance Index (CPI) is calculated:

$$PF_3 = \frac{CumulativeBCWP(latest) - CumulativeBCWP(3 \cdot months \cdot ago)}{CumulativeACWP(latest) - CumulativeACWP(3 \cdot months \cdot ago)} \tag{2.17}$$

where BCWP= Budgeted Cost of Work Performed.

For the Six-Month Average method the six-month CPI is calculated:

$$PF_6 = \frac{Cumulative \cdot BCWP(latest) - Cumulative \cdot BCWP(6 \cdot months \cdot ago)}{Cumulative \cdot ACWP(latest) - Cumulative \cdot ACWP(6 \cdot months \cdot ago)} \tag{2.18}$$

for the Cumulative CPI method, the Cumulative CPI is used:

$$PF_{cum} = Cumulative \cdot CPI = \frac{Cumulative \cdot BCWP(latest)}{Cumulative \cdot ACWP(latest)} \tag{2.19}$$

for the current CPI method, the current period CPI is used:

$$PF_{cur} = Current \cdot Period \cdot CPI \cdot = \frac{Current \cdot Period \cdot BCWP}{Current \cdot Period \cdot ACWP} \tag{2.20}$$

for the weighted Cost-Schedule method, both the Cumulative CPI and Schedule Performance Index (SPI) are used:

$$PF_{cs} = A \times Cumulative \cdot CPI + B \times Cumulative \cdot SPI \tag{2.21}$$

for the Performance Factor method:

$PF_p = Value \cdot chosen \cdot by \cdot the \cdot analyst$, such that:

$$CAC = BCWR \times PF_p + CumulativeACWP \tag{2.22}$$

3.0 COMPARING THE MODELS

The use of exponential smoothing and time series models needs some basic ground rules for them to be acceptable forecasters. The largest drawback of using these methods is that more data is required to obtain good models than for the models of PA. A reasonable suggestion is not to use these techniques if a contract duration is less than thirty-six months. In addition, it is wise to use as much data as possible to *fit* the models. Therefore, it is further suggested that they be used when at least 50% of the contract has been performed.

Since these models reasonably track anomalies, they are well suited for use on contracts that are having difficulty because of underestimating the amount of resources and time required to complete work packages, delays in subcontractor deliveries, and not programming enough *risk* into historically difficult work packages such as software development and software integration and testing. The data obtained for this study are from two contracts that had a history of anomalous cost efficiency. The durations of these contracts are thirty-seven and forty-four months. They are identified herein as R-37 and T-44, respectively. The exponential smoothing and time series methods were compared to several PA methods for approximately 50% to 80% of the contract work performed.

3.1 Exponential Smoothing Models

The first of the results are for the T-37 data where eighteen data points (cumulative ACWP) were used to fit the model. Six-month increments were used to obtain a perspective of how good the models are and when they can be best applied.

Figures 3.1, 3.2, and 3.3 illustrate simulations of the original data using linear, quadratic, and sine with linear trend. These seem to be the most logical of the available exponential smoothing models for fitting the CPR data. The simulations indicate that the linear exponential smoothing model best fits the data. The other simulations are shown for comparison.

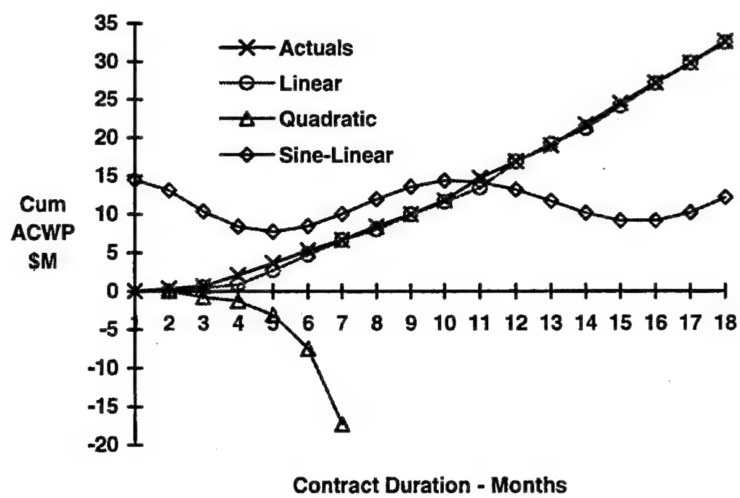

Figure 3.1 Simulated R-37 Data vs. Actual Cumulative ACWP for Eighteen Data Points

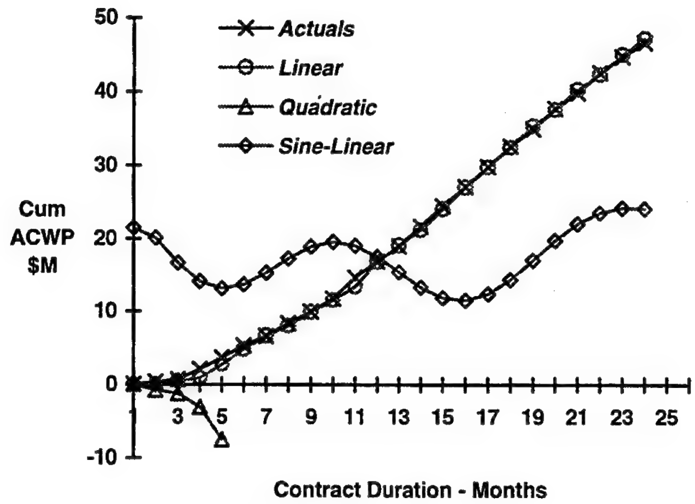

Figure 3.2 Simulated R-37 Data vs. Actual Cumulative ACWP for Twenty-Four Data Points

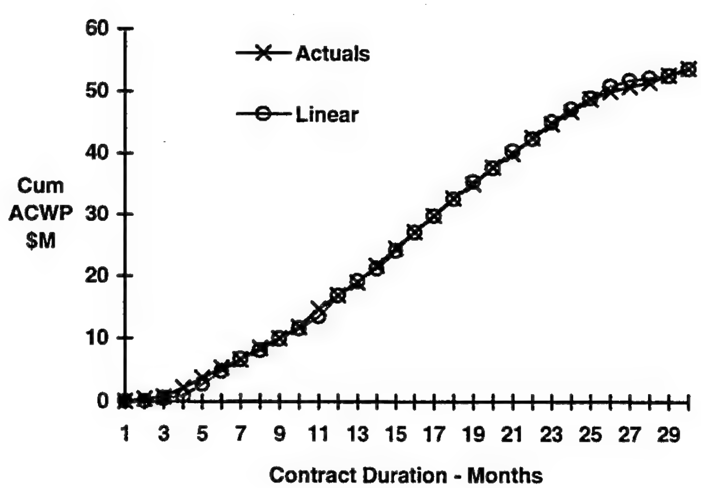

Figure 3.3 Simulated R-37 Data vs. Actual Cumulative ACWP for Thirty Data Points

The same general results were obtained from the T-44 CPR data. In this case, the Linear Smoothing model gave the best simulations. Figures 3.4, 3.5 and 3.6 illustrate the best model fit for the data.

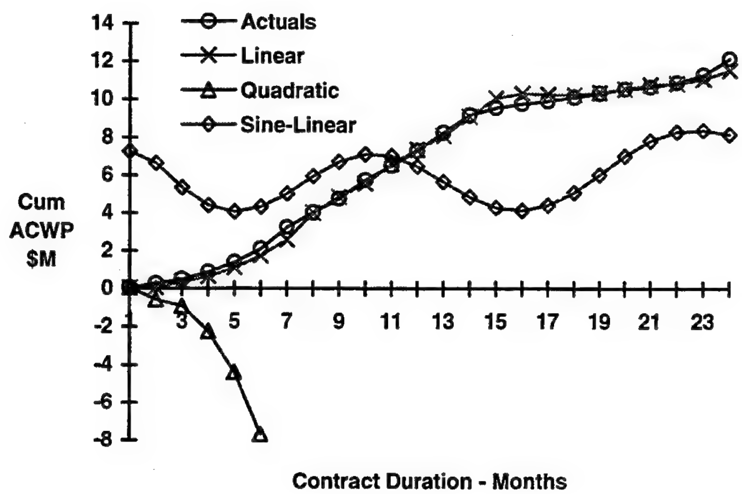

Figure 3.4 Simulated T-44 Data vs. Actual Cumulative ACWP for Twenty-Four Data Points

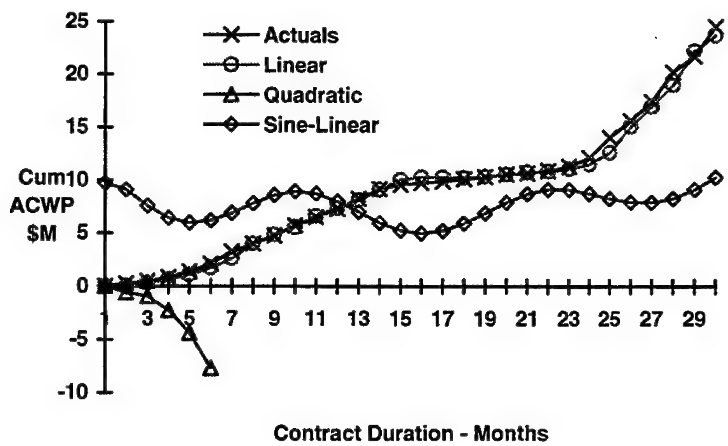

Figure 3.5 Simulated T-44 Data vs. Actual Cumulative ACWP for Thirty Data Points

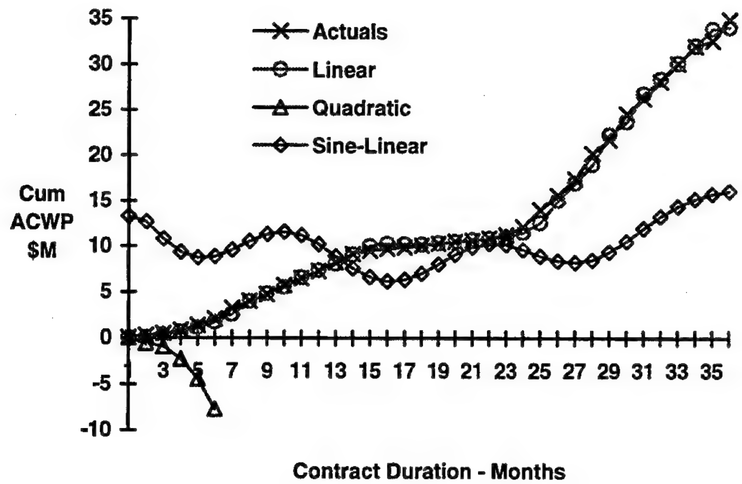

Figure 3.6 Simulated T-44 Data vs. Actual Cumulative ACWP for Thirty-Six Data Points

These results strongly indicate that the linear exponential smoothing model provides the best fit for the R-37 and T-44 data and that the model is useful for forecasting estimates-at-completion when contracts are at least 50% complete (with 18 - 24 data points). Applying this method to contracts that are less than 50% complete or for contracts of short duration (e.g., less than 36 months) is not recommended.

3.2 Time Series Models

Time series models are more difficult to fit than exponential smoothing models for the R-37 and T-44 data. They require a larger data base to obtain adequate models for forecasting. Though the simulations are not as good as the exponential smoothing models, they do have the ability to update forecasts as new data become available. Figures 3.7 through 3.12 clearly indicate that as more information becomes available, the better the model *fit*. In the case of the R-37 data, the best fit is for 30 data points (approximately 80% complete) with potential *overestimates* for the model fit with 24 data points (approximately 65% complete). In somewhat of a contrast, the fit for the T-44 data was excellent for all models starting at 24 data points (approximately 50% complete). These results reinforce *when* to use the time series approach. First, contracts of less than 36 months duration should not be considered prime candidates for this method. Second, at least 50% (or n>24 data points) of the contract should be complete, and, third, it is probably best to use the time series approach for *problem children* (e.g., when there is a wide range of cost efficiency, underestimated risk, over-target baselines) since this method has an updating capability [1], [3].

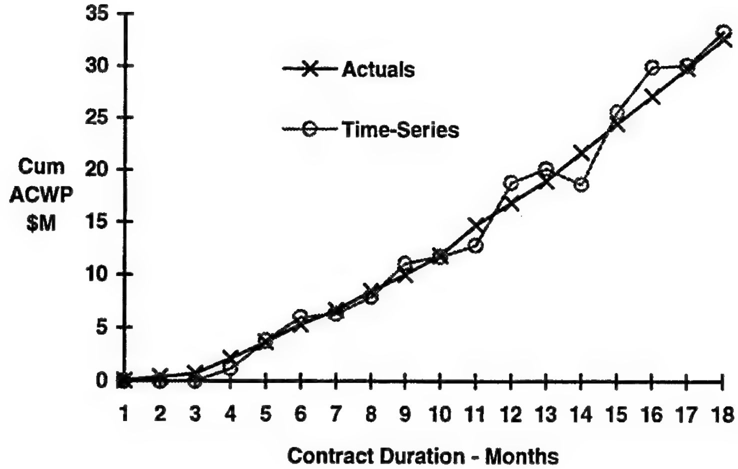

Figure 3.7 Simulated R-37 Data vs. Actual Cumulative ACWP for Eighteen Data points

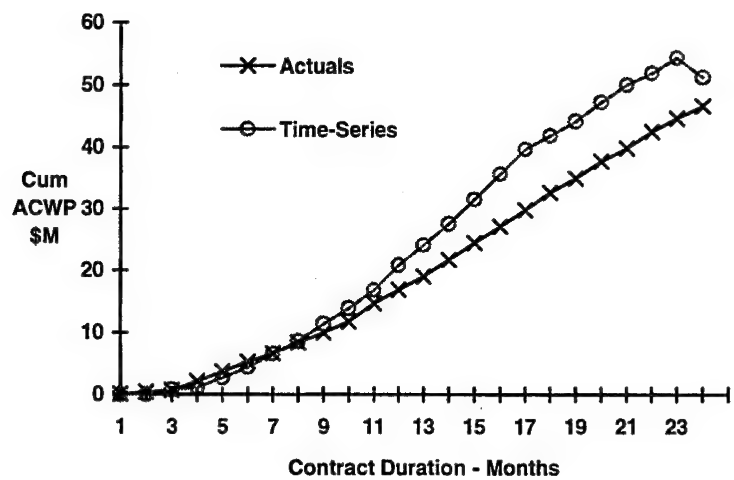

Figure 3.8 Simulated R-37 Data vs. Actual Cumulative ACWP for Twenty-Four Data Points

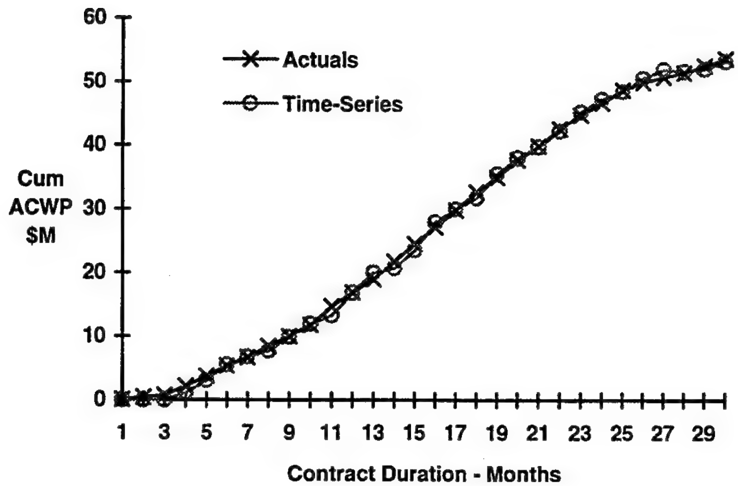

Figure 3.9 Simulated R-37 Data vs. Actual Cumulative ACWP for Thirty Data Points

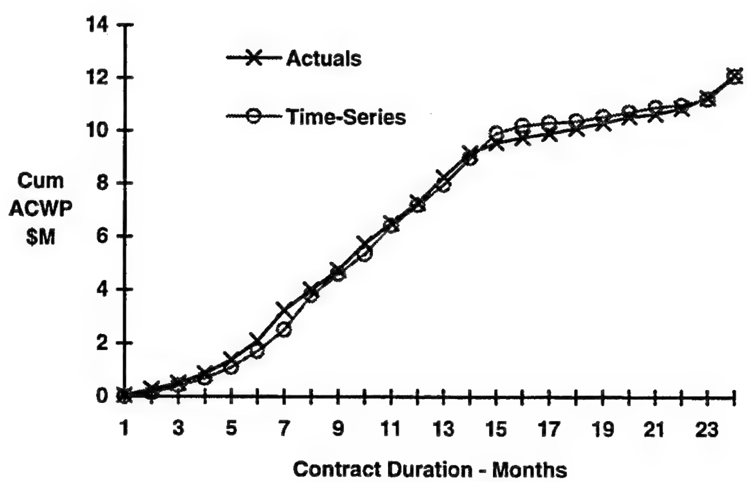

Figure 3.10 Simulated T-44 Data vs. Actual Cumulative ACWP for Twenty-Four Data Points

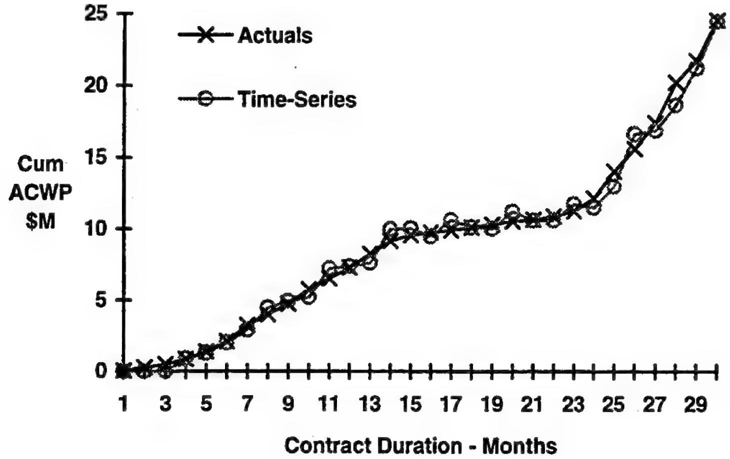

Figure 3.11 Simulated T-44 Data vs. Actual Cumulative ACWP for Thirty Data Points

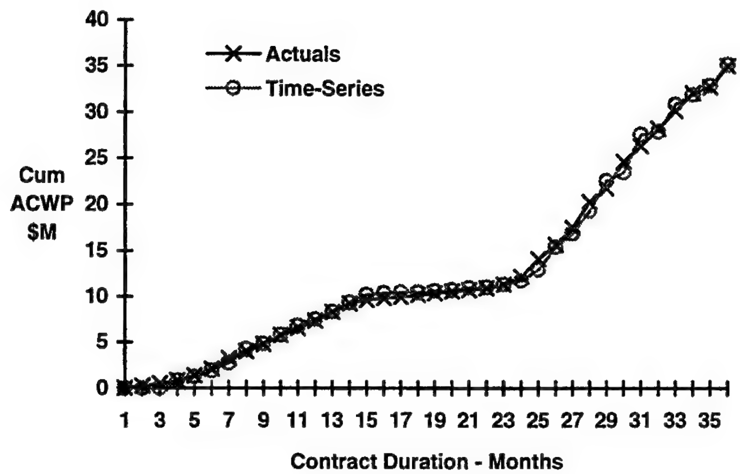

Figure 3.12 Simulated T-44 Data vs. Actual Cumulative ACWP for Thirty-Six Data points

4.0 THE FITTED MODELS

This section contains the parametric equations that were obtained for the R-37 and T-44 data. The parameters for the exponential smoothing models were based on the *history* of the data and the most recent smoothed data for each six-month interval (e.g., 18 months of data, 24 months of data). The smoothed parameters are shown in Table 4.1.

Table 4.1 Estimated Parameters for the R-37 and T-44 Exponential Smoothing Models

Data Source	No. Data Points	Parameter A_1	Parameter A_2
R-37	18	32.66	2.75
	24	46.78	2.17
	30	53.74	1.05
T-44	24	12.08	0.55
	30	24.49	2.31
	36	34.88	1.71

Using these parameters, the following equations were obtained for the data:

For the R-37 data:

$$\xi(t+\tau) = 32.66 + 2.75t \text{ at } x_{18} \tag{4.1}$$

$$\xi(t+\tau) = 46.78 + 2.17t \text{ at } x_{24} \tag{4.2}$$

$$\xi(t+\tau) = 53.74 + 1.05t \text{ at } x_{30} \tag{4.3}$$

For the T-44 data:

$$\xi(t+\tau) = 12.08 + 0.55t \text{ at } x_{24} \tag{4.4}$$

$$\xi(t+\tau) = 24.49 + 2.31t \text{ at } x_{30} \tag{4.5}$$

$$\xi(t+\tau) = 34.88 + 1.71t \text{ at } x_{36} \tag{4.6}$$

The parameters for the time series approach including the order of the model and the order of the difference filter are shown in Table 4.2.

The order of the model and filter are in the following format:

ARIMA(m,p,q)

where m is the order of the autoregressive component of the model, p is the order of the filter, and q is the order of the moving average component of the model. ARIMA is the acronym for the mixed Autoregressive Integrated Moving Average model. All characterizations of the data required filtering and were mixed models (see equation 2.5).

Table 4.2 Estimated Parameters for the R-37 and T-44 Time Series Models

Data Source	Data Points	Model Order	AR α_1	AR α_2	MA β_1	MA β_2	MA β_3	Model Mean, μ
R-37	18	(1,2,2)	-.1094	-	.7031	-.6875	-	.1568
	24	(2,2,3)	-.2813	.2500	.4375	-.6563	.4844	.2010
	30	(2,1,3)	.6719	.2969	.2500	-.4219	.4688	.1851
T-44	24	(1,1,1)	.9375	-	.2969	-	-	.5260
	30	(1,2,3)	-.5469	-	.5469	-.3438	.5313	.0931
	36	(1,2,2)	-.2969	-	-.4219	.5469	-	.0617

Using the corresponding parameters, the following difference equations were obtained for the filtered R-37 series:

for 18 data points:

$$(w_t-.1568)=-.1094(w_{t-1}-.1568)+Z_t+.7031Z_{t-1}-.6875Z_{t-2} \tag{4.7}$$

for 24 data points:

$$(w_t-.2010)=-.2813(w_{t-1}-.2010)+.2500(w_{t-2}-.2010)+Z_t+.4375Z_{t-1}-.6563Z_{t-2}+.4844Z_{t-3} \tag{4.8}$$

and for 30 points:

$$(y_t-.1851)=.6719(y_{t-1}-.1851)+.2969(y_{t-2}-.1851)+Z_t+.2500Z_{t-1}-.4219Z_{t-2}+.4688Z_{t-3} \tag{4.9}$$

For the T-44 filtered series, the difference equations are:
for 24 data points:

$$(y_t-.5260)=.9375(y_{t-1}-.5260)+Z_t+.2969Z_{t-1} \tag{4.10}$$

for 30 data points:

$$(w_t - .0931) = -.5469(w_t - .0931) + Z_t + .5469t - 3Z_{t-1} - .3438Z_{t-2} + .5313Z_{t-3} \quad (4.11)$$

and for 36 data points:

$$(w_t - .0617) = -.2969(w_{t-1} - .0617) + Z_t - .4219Z_{t-1} + .5469Z_{t-2} \quad (4.12)$$

Since the data required filtering, *backwards* filters must be inserted into the difference equations to obtain the appropriate forecasting models that characterize the R-37 and T-44 data. Using equations 2.1 and 2.2, the following forecasting models were obtained:

for the R-37 data, 18 data points:

$$\hat{x}_t = 1.8906x_{t-1} - .7813x_{t-2} - .1094x_{t-3} + Z + .7031Z_{t-1} - .6875Z_{t-2} + .1740 \quad (4.13)$$

for 24 data points:

$$\hat{x}_t = 1.7188x_{t-1} - .1875x_{t-2} - .7813x_{t-3} + 2500x_{t-4} + Z_t + .4375Z_{t-1} - .6563Z_{t-2} + .4844Z_{t-3} + .1441 \quad (4.14)$$

and for 30 data points:

$$\hat{x}_t = 1.6719x_{t-1} - .3750x_{t-2} - .2969x_{t-3} + Z_t - .4219Z_{t-1} + .5469Z_{t-2} + .0329 \quad (4.15)$$

And, for the T-44 data, for 24 data points:

$$\hat{x}_t = 1.9375x_{t-1} - .9375x_{t-2} + Z_t + .2969Z_{t-1} + .0329 \quad (4.16)$$

for 30 data points:

$$\hat{x}_t = 1.4531x_{t-1} + .0938x_{t-2} - .5469x_{t-3} + Z_t + .5469Z_{t-1} - .3438Z_{t-2} + .5313Z_{t-3} + .1441 \quad (4.17)$$

and finally, for 36 data points:

$$\hat{x}_t = 1.7031x_{t-1} - .4063x_{t-2} - .2969x_{t-3} + Z_t - .4219Z_{t-1} + .5469Z_{t-2} + .0799 \quad (4.18)$$

Equations 4.13 through 4.18 were used to obtain the simulated series for each six-month interval and to forecast future values of the time series.

5.0 RESULTS AND CONCLUSIONS

In previous sections a comparison was made between the exponential smoothing and time series models. This was done on the basis of fitting models to actual data and then using them to simulate the original data series. This approach led to conclusions on *when* to use these models and how much data would be needed for a good *fit*. In this section, the models of section 4.0 are used to forecast Estimates-at-Completion (EAC) for the R-37 and T-44 series. Table 5.1 illustrates the differences between exponential smoothing and time series forecasts for the T-37 data relative to those obtained from several methods found in Performance Analyzer.

Table 5.1 Comparison of Forecasts for the R-37 Data
($M)

Data Points	*LRE*	*Time Series*	*Exponential Smoothing*	*3 Month Average*	*6 Month Average*	*Cum CPI*	*Cost-Schedule*	*Perform. Factor*
18	46.78	64.97	73.90(L)	62.39	60.08	55.91	55.20	51.70
24	54.68	59.30	57.08(Q)	53.01	56.00	56.23	56.17	55.98
30	56.43	59.99	61.01	59.99	57.07	56.76	56.23	56.58

The current LRE is $59.94M and the contractor's performance is on track.

For 18 data points (approximately 50% complete), all forecasts indicate that the LRE is understated. At this stage of the contract, picking the best forecast is purely judgmental. For instance, the six-month average forecast is closest to the *current LRE.* However, there is almost a 25% difference relative to the LRE for that time frame. As the number of months of data increases to 24, the contract is approximately 67% complete and the time series forecast is within $0.64M of the *current LRE.* This is the best of all the forecasts. With 30 months of data available, the contract is approximately 80% complete and the time series forecast is within $0.06M of the *current LRE* as is the three-month average forecast. The exponential smoothing forecast is also very good being within $0.07M of the *current LRE.* The advantage of the time series approach is the accuracy that can be achieved much earlier than with the other methods. The exponential smoothing model overestimated the *current LRE* but got better as more data became available. Generally, the PA methods underestimated the *current LRE.*

Table 5.2 illustrates the model differences for the T-44 data.

Table 5.2 Comparison of forecasts for the T-44 Data
($M)

Data Points	*LRE*	*Time Series*	*Exponential Smoothing*	*3 Month Average*	*6 Month Average*	*Cum CPI*	*Cost-Schedule*	*Perform. Factor*
24	40.29	43.84	23.12	37.91	40.43	42.77	42.56	40.18
30	43.67	55.24	56.84	47.05	44.74	45.06	45.47	43.52
36	45.36	46.08	48.59	47.90	46.52	47.45	47.59	46.63

The current LRE is $46.10M with the contract 95% complete, and the contract is on track.

For less than 24 months of data, the T-44 series PA forecasts underestimated the current LRE as did the time series and exponential smoothing models. However, as more data became available forecasts got better.

At 24 months complete, the time series forecast was better overall. At 30 months complete, the PA methods were generally better relative to the current LRE. With 36 months of data available, all models were within 5% of the current LRE with the time series closest with a difference of less than 1%.

Better estimates were expected for this series for 30 data points. It should be noted, however, due to funding constraints, a work slowdown was ordered when the contract was about 65% complete. After funding was restored and the contractor's cost efficiency was restored, the forecasts were better. For 36 months of data available, the time series forecast was within $0.02M of the current LRE. The other models gave overestimated EACs.

These results indicate that the time series approach provides better EACs much earlier than exponential smoothing and the methods of PA. Several conclusions can be drawn from the results:

a. Time series models provide better forecasts earlier than the other methods used.

b. Exponential smoothing models, which are easier to compute, gave better EACs as more data became available.

c. Performance Analyzer methods generally underestimated the EACs for the series early on.

d. It is not recommended that time series models be used for contracts of short duration, e.g., less than 36 months.

e. The time series approach needs at least 24 months of data to provide adequate forecasting models.

From the analyst's point-of-view, the results of this study may not be the same for other CPR data. Confidence in forecasts that are used for managing contracts can only be attained from frequent use of the available forecasting methods. For some methods, more monthly data are required to obtain reliable EACs. For instance, the three-month average method requires three months of data in contrast to time series and exponential smoothing which require at least 24 months of data. In the long run, the analyst's judgment based on experience is the determining factor in choosing the best forecasting model.

REFERENCES

1. Box, G.E.P, and G.M. Jenkins, Time Series Analysis, Forecasting, and Control. San Francisco, California: Holden-Day 1970.

2. Leavenbach, H., and J.P. Cleary, The Professional Forecaster. Belmont, California: Lifetime Learning Publications, 1982

3. D'Accardi, R.J., The Use of Time Series Forecasting in Contractor Performance Analysis. CECOM Research and Development Technical Report, CECOM-TR-94-12, US Army Communications-Electronics Command, Directorate of Resource Management, Fort Monmouth NJ, November 1994.

4. Kendall, M.G., and A. Stuart, The Advanced Theory of Statistics, Volume 3. London, England: Griffen, 1966.

5. Brown, R.G., Smoothing, Forecasting and Prediction of Discrete Time Series. Toronto, Canada. Prentice-Hall, 1963.